

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平8-69446

(43)公開日 平成8年(1996)3月12日

(51)Int.Cl.<sup>8</sup>

G 0 6 F 15/18  
17/00

識別記号

5 5 0 Z 8837-5L

9069-5L

庁内整理番号

F I

G 0 6 F 15/ 20

技術表示箇所

F

審査請求 有 請求項の数 9 O L (全 8 頁)

(21)出願番号 特願平6-203547

(22)出願日 平成6年(1994)8月29日

(71)出願人 000004237

日本電気株式会社  
東京都港区芝五丁目7番1号

(72)発明者 馬見塚 拓

東京都港区芝五丁目7番1号 日本電気株  
式会社内

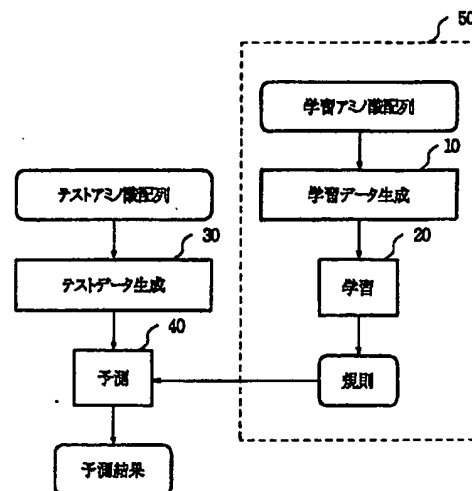
(74)代理人 弁理士 京本 直樹 (外2名)

(54)【発明の名称】 タンパク質立体構造予測方法

(57)【要約】

【目的】 システインを含む立体構造未知のテストアミノ酸配列が与えられたときに、S S結合を構成するようなシステインの組み合わせを高い精度で予測する。

【構成】 学習データ生成ステップ30では、立体構造既知のタンパク質アミノ酸配列から、立体構造を予測するために、学習データを抽出する。学習ステップ20では、学習データから確率的規則の学習を行う。テストデータ生成ステップ10では、与えられたテストアミノ酸配列に対し、テストデータの生成を行う。予測ステップ40では、生成されたテストデータおよび学習により得られた確率的規則を用いて予測を行う。



## 【特許請求の範囲】

【請求項1】 立体構造既知のタンパク質アミノ酸配列から、立体構造を予測するために、学習データを抽出する学習データ生成ステップと、学習データから確率的規則の学習を行う学習ステップと、与えられたテストアミノ酸配列に対し、テストデータの生成を行うテストデータ生成ステップと、生成されたテストデータおよび学習により得られた確率的規則を用いて予測を行う予測ステップとを含むことを特徴とするタンパク質立体構造予測方法。

【請求項2】 前記学習データ生成ステップが、対象とする化学結合を含む立体構造既知のタンパク質アミノ酸配列に対して、相同性を有するアミノ酸配列のアライメントを行ない、化学結合を構成するアミノ酸位置に相当するアミノ酸を各位置の学習データとして抽出することを特徴とする請求項1記載のタンパク質立体構造予測方法。

【請求項3】 前記学習ステップが、前記学習データのアミノ酸位置の組み合わせにおけるアミノ酸種類の変化から、確率的規則を用いることにより、この確率的規則の実数値パラメータを推定することを特徴とする請求項1または請求項2記載のタンパク質立体構造予測方法。

【請求項4】 前記テストデータ生成ステップが、与えられた立体構造未知のテストアミノ酸配列に対して、相同性を有するアミノ酸のアライメントを行ない、対象とする化学結合を構成するアミノ酸位置に相当するアミノ酸を各位置のテストデータとして抽出することを特徴とする請求項1記載のタンパク質立体構造予測方法。

【請求項5】 前記予測ステップが、前記テストデータ生成ステップにより生成された各アミノ酸位置のテストデータに対して、前記学習ステップにより学習されたアミノ酸種類の変化の確率的規則を使用し、各位置の組み合わせの尤度を計算するステップと、計算された尤度を用いて最適な化学結合の組み合わせを選出するステップとを含むことを特徴とする請求項1または請求項4記載のタンパク質立体構造予測方法。

【請求項6】 前記学習データ生成ステップが、SS結合を含む立体構造既知のタンパク質アミノ酸配列に対して、相同性を有するアミノ酸配列のアライメントを行ない、SS結合を構成するシステイン位置に相当するアミノ酸を各位置の学習データとして抽出することを特徴とする請求項1記載のタンパク質立体構造予測方法。

【請求項7】 前記学習ステップが、前記学習データのシステイン位置の組み合わせにおけるアミノ酸種類の変化から、確率的規則を用いることにより、この確率的規則の実数値パラメータを推定することを特徴とする請求項1または請求項6記載のタンパク質立体構造予測方法。

【請求項8】 前記テストデータ生成ステップが、与えられた立体構造未知のテストアミノ酸配列に対して、相

同性を有するアミノ酸配列のアライメントを行ない、システイン位置に相当するアミノ酸を各位置のテストデータとして抽出することを特徴とする請求項1記載のタンパク質立体構造予測方法。

【請求項9】 前記予測ステップが、前記テストデータ生成ステップにより生成された各システイン位置のテストデータに対して、前記学習ステップにより学習されたアミノ酸種類の変化の確率的氣息を使用し、各システイン位置の組み合わせの尤度を計算するステップと、計算された尤度を用いて最適なSS結合の組み合わせを選出するステップとを含むことを特徴とする請求項1または請求項8記載のタンパク質立体構造予測方法。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は、立体構造未知のタンパク質アミノ酸配列から、そのタンパク質の立体構造を予測する方法に関する。

## 【0002】

【従来の技術】 タンパク質は、20種類のアミノ酸の一次的な結合により形成されており、その一次的な配列が折り畳まれた立体構造として存在する。

【0003】 タンパク質の立体構造は、X線結晶回折などの物理化学的手法を用いることにより決定可能であるが、このためには、数億円もする機器と、数カ月にも及ぶ時間、また、優れた解析技術を持った専門家などが必要であり、容易ではない。

【0004】 タンパク質の構造はその機能と密接に関係しており、生体内のほぼすべての活動を司るタンパク質の立体構造を、計算機を用いて短時間に容易に予測することができれば、分子生物学をはじめとして、生化学、医学、薬学、さらには農学など数多くの分野に大きな大きな寄与ができると考えられている。

【0005】 現在、計算機により、高い精度で立体構造予測可能な方法としては、ブランデル (T.L.Blundell) らによるホモロジー法 (T.L.Blundell, B.L.Sibanda, M.J.E.Sternberg, and J.M.Thornton (1987) Knowledge-based Prediction of Protein Structures and the Design of novel Molecules, Nature, vol. 326:347-352) やボウイ (J.U.Bowie) らによるインバースフォールディング法 (J.U.Bowie, R.Luthy, and D.Eisenberg (1991) A Method to Identify Protein Sequences that Fold into a Known Three-Dimensional Structure, Science, vol. 253:164-170) などがあるが、これらは、与えられたテストタンパク質が、データベース中の既に立体構造が決定されたタンパク質のいずれかと高い相同性を有することが条件となっている。

【0006】 しかし、現在立体構造データベースに含まれているタンパク質数は、数百程度にすぎず、数億あるいはそれ以上とも言われる地球上のタンパク質数に較べれば、はるかに少ない。したがって、計算機を用いた前

述の方法では、与えられたテストタンパク質と相溶性の固い立体構造既知のタンパク質を見つけることが非常に難しく、有効に働かない場合の方が多い。

【0007】そのため、タンパク質全体の構造を一気に決めるのではなく、立体構造の部分的で特徴的な部分を見いだす方法が考えられている。

【0008】この場合、最も代表的な部分構造としては、二次構造が挙げられる。タンパク質の二次構造とは、タンパク質の立体構造の中で出現するままとりのある局所構造を指し、 $\alpha$ ヘリックスや $\beta$ シートと呼ばれるものが代表的である。与えられたテストタンパク質に対し、その二次構造を予測する方法としては、例えば、ロスト (B.Rost) らによるマルチプルアライメントとニューラルネットワークを用いた手法 (B.Rost, and C.Sander(1993) Prediction of Protein Secondary Structure a t Better Than 70 % Accuracy, J.Mol.Biol.vol.232:548-599) が知られているが、最も優れたこの手法においても、予測率は70%程度にとどまっており、実用に耐えない。

【0009】ここで、配列上では遠距離に位置する2つのアミノ酸が、立体構造において近接し、化学結合を構成する場合を考えてみよう。例えば、立体構造上、近接するシステインの硫黄 (S) 原子間にはSS結合 (disulfide bond) が形成される場合がある。SS結合を構成できる残基は、システインのみなので、SS結合は2つのシステイン同士によってしか形成されない。したがって、システイン原子を複数含み、かつSS結合を構成する構造未知のタンパク質アミノ酸配列が与えられた場合に、どのシステインの組み合わせがSS結合を構成するかが高い精度でわかれば、立体構造を予測する上での重要な手がかりを与える。

【0010】しかし、今まで、システインを含む立体構造未知のタンパク質アミノ酸配列が与えられた場合に、計算機を用いて、どのシステイン同士の組み合わせがSS結合を構成するかを解決する方法は、全く考えられなかった。また、SS結合に限らず、計算機を用いて、遠距離に位置するアミノ酸同士の化学結合の組み合わせを同定する方法は、全く考えられなかった。

【0011】

【発明が解決しようとする課題】SS結合を初めとしたアミノ酸同士の遠距離相互作用は、立体構造をより安定化するために重要であるが、そのような化学結合の形成が立体構造への折り畳みを進める機動力になるわけではない。つまり、立体構造未知のタンパク質アミノ酸配列において、SS結合を構成するようなシステインの組み合わせは、それらの周辺残基の影響と無関係に決定すると考えられる。したがって、与えられた1本のアミノ酸配列から、その中に含まれるシステインの組み合わせを決定することは困難だった。

【0012】そこで、本発明の目的は、与えられた一本

のアミノ酸配列データのみならず、SS結合を初めとした化学結合の組み合わせを検出するために有効なデータをアミノ酸配列データベースから生成する方法を提供するとともに、生成したデータを使用して、SS結合をはじめとした化学結合の組み合わせを予測する方法を提供することにある。

【0013】

【課題を解決するための手段】第1の発明のタンパク質立体構造予測方法は、立体構造既知のタンパク質アミノ酸配列から、立体構造を予測するために、学習データを抽出する学習データ生成ステップと、学習データから確率的規則の学習を行う学習ステップと、与えられたテストアミノ酸配列に対し、テストデータの生成を行うテストデータ生成ステップと、生成されたテストデータおよび学習により得られた確率的規則を用いて予測を行う予測ステップとを含むことを特徴とする。

【0014】第2の発明のタンパク質立体構造予測方法は、第1の発明において、前記学習データ生成ステップが、対象とする化学結合を含む立体構造既知のタンパク質アミノ酸配列に対して、相溶性を有するアミノ酸配列のアライメントを行ない、化学結合を構成するアミノ酸位置に相当するアミノ酸を各位置の学習データとして抽出することを特徴とする。

【0015】第3の発明のタンパク質立体構造予測方法は、第1または第2の発明において、前記学習ステップが、前記学習データのアミノ酸位置の組み合わせにおけるアミノ酸種類の変化から、確率的規則を用いることにより、この確率的規則の実数値パラメータを推定することを特徴とする。

【0016】第4の発明のタンパク質立体構造予測方法は、第1の発明において、前記テストデータ生成ステップが、与えられた立体構造未知のテストアミノ酸配列に対して、相溶性を有するアミノ酸のアライメントを行ない、対象とする化学結合を構成するアミノ酸位置に相当するアミノ酸を各位置のテストデータとして抽出することを特徴とする。

【0017】第5の発明のタンパク質立体構造予測方法は、第1または第4の発明において、前記予測ステップが、前記テストデータ生成ステップにより生成された各アミノ酸位置のテストデータに対して、前記学習ステップにより学習されたアミノ酸種類の変化の確率的規則を使用し、各位置の組み合わせの尤度を計算するステップと、計算された尤度を用いて最適な化学結合の組み合わせを選出するステップとを含むことを特徴とする。

【0018】第6の発明のタンパク質立体構造予測方法は、第1の発明において、前記学習データ生成ステップが、SS結合を含む立体構造既知のタンパク質アミノ酸配列に対して、相溶性を有するアミノ酸配列のアライメントを行ない、SS結合を構成するシステイン位置に相当するアミノ酸を各位置の学習データとして抽出するこ

とを特徴とする。

【0019】第7の発明のタンパク質立体構造予測方法は、第1または第6の発明において、前記学習ステップが、前記学習データのシステイン位置の組み合わせにおけるアミノ酸種類の変化から、確率的規則を用いることにより、この確率的規則の実数値パラメータを推定することを特徴とする。

【0020】第8の発明のタンパク質立体構造予測方法は、第1の発明において、前記テストデータ生成ステップが、与えられた立体構造未知のテストアミノ酸配列に対して、相同性を有するアミノ酸配列のアライメントを行ない、システイン位置に相当するアミノ酸を各位置のテストデータとして抽出することを特徴とする。

【0021】第9の発明のタンパク質立体構造予測方法は、第1または第8の発明において、前記予測ステップが、前記テストデータ生成ステップにより生成された各システイン位置のテストデータに対して、前記学習ステップにより学習されたアミノ酸種類の変化の確率的気息を使用し、各システイン位置の組み合わせの尤度を計算するステップと、計算された尤度を用いて最適なSS結合の組み合わせを選出するステップとを含むことを特徴とする。

【0022】

【実施例】次に、本発明について図面を参照して説明する。

【0023】図1は、本発明によるタンパク質立体構造予測方法の一実施例を示すフローチャートである。この実施例では、対象とする化学結合として、システイン同士の結合からなるSS結合を扱うものとする。

【0024】ステップ10は、第2および第6の発明に対応する。図2は、このステップでの手順の模式図である。このステップでは、SS結合を含む学習アミノ酸配列を入力とする。この配列に対して相同性を有する配列をアミノ酸配列データベースから抽出し、相同性の高い順番に並べる。それらのアライメントを行ない、学習アミノ酸配列のSS結合を構成するシステイン位置に相当するアミノ酸を各システイン位置の学習データとする。ここで、相同性を有する配列とは、具体的には、例えば、50%から30%の相同性を有する配列などが挙げられる。

【0025】ステップ20は、第3および第7の発明に対応する。このステップでは、ステップ10で生成した学習データを入力とし、確率的規則を用いることにより、この確率的規則の実数値パラメータを推定する。ステップ20での確率的規則の構造を以下に示す。

【0026】ここでの確率的規則とは、2つのシステイン位置のアライメントにより得られたアミノ酸データに対して、それらの位置がSS結合を構成する確率を与える確率分布のことである。

【0027】配列全体におけるシステイン位置数を  $n$

、アライメントされた配列データ数を  $m_L$ 、さらに、2つのシステイン位置の組み合わせに番号を付け、 $i$  番目の組み合わせを  $C_i [i=1, \dots, n_L(n_L-1)/2]$  とする。ここで、 $i$  番目の組み合わせがSS結合 (S) を構成する確率 ( $S | C_i$ ) が次のように計算できると仮定する。

【0028】

$$P(S | C_i) = \prod_{j=1}^{m_L-1} P(S | v_{ij}) \quad \left( i=1, \dots, \frac{n_L(n_L-1)}{2} \right)$$

【0029】ただし、 $v_{ij}$  は、 $i$  番目の組み合わせを構成する2つのシステイン位置のアライメントデータにおいて、となり合う  $j$  ( $j=1, \dots, m_L-1$ ) 番目と  $j+1$  番目のデータのアミノ酸種類の変化を表す。さらに、この変化に対して、SS結合が構成される確率を  $P(S | v_{ij})$  とする。

【0030】 $v_{ij}$  の具体的な表現には、例えば、次のようなものがある。

【0031】まず、アミノ酸種類を「システインである (以下、Cとする)」、「システインでない (以下、Nとする)」のいずれかに分ける。この場合、2つの位置に対しては、「ともにシステイン (CC)」、「システインとシステインでない (CN)」、「システインでない」とシステイン (NC)」、「ともにシステインでない (NN)」の4種類がある。さらに、2つの位置の  $j$  番目のデータから  $j+1$  番目のデータへのアミノ酸種類の変化は、表1に示す16種類の変化を考えることができる。

【0032】

【表1】

番号	組み合わせ位置の変化
1	CC→CC
2	CC→CN
3	CC→NC
4	CC→NN
5	CN→CC
6	CN→CN
7	CN→NC
8	CN→NN
9	NC→CC
10	NC→CN
11	NC→NC
12	NC→NN
13	NN→CC
14	NN→CN
15	NN→NC
16	NN→NN

【0033】さらに、表1の番号に対応する  $k$  ( $k=1, \dots, 16$ ) での確率パラメータ  $p(k)$  を用いて、 $v_{ij}=k$  であるとき、 $P(S | v_{ij}) = p(k)$  とす

る。

【0034】確率パラメータは、ステップ10で得られた学習データを用いて推定する。表1の16種類のアミノ酸変化において、 $N_k(i)$  ( $k=1, \dots, 16$ ) を  $i$  番目のシステイン位置の組み合わせで  $k$  番目のアミノ酸変化が得られた個数とする。ここで、学習データのシステイン位置の組み合わせの中で、SS結合を構成する組み合わせの集合を  $S$ 、構成しない組み合わせの集合を

$$\bar{S}$$

【0036】とした場合に、 $k$  番目のアミノ酸変化の正例数

【0037】

$$N_k^+$$

【0038】 $k$  番目のアミノ酸変化の負例数

【0039】

$$N_k^-$$

【0040】を次式で求める。

【0041】

$$N_k^+ = \sum_{i \in S} N_k(i) \quad (k=1, \dots, 16)$$

$$N_k^- = \sum_{i \in \bar{S}} N_k(i) \quad (k=1, \dots, 16)$$

【0042】 $k$  番目のアミノ酸変化における推定値を

【0043】

$$\hat{p}(k)$$

【0044】とし、例えば、次式のラプラス推定量により計算する。

【0045】

$$\hat{p}(k) = \frac{N_k^+ + 1}{N_k^+ + N_k^- + 2} \quad (k=1, \dots, 16)$$

【0046】ただし、推定量はラプラス推定量のみならず、多くの推定量が使用できる。

【0047】ステップ30は、第4および第5の発明に対応する。このステップでの手順も図2に示されている。このステップでは、ステップ10と同様に、与えられたテストアミノ酸配列に対し、相同性を有する配列を配列データベースから抽出し、相同性の高い順番に並べる。それらのアライメントを行ない、システイン位置に相当するアミノ酸をテストデータとする。ただし、ここで与えられるテストアミノ酸配列とは、SS結合を構成し、SS結合数が実験的にわかっているとする。また、相同性を有する配列とは、具体的には、例えば、50%から30%程度の相同性を有する配列などが考えられ

る。

【0048】ステップ40は、第5および第9の発明に対応する。このステップでは、ステップ20で学習された確率的規則を使用し、ステップ30で生成されたシステイン位置のテストデータの各組み合わせに対して、その活性度を計算する。さらに、活性度から、SS結合の組み合わせを予測する。

【0049】図3は、ステップ40の手順を詳細に示すフローチャートである。

10 【0050】まず、ステップ60での活性度の計算方法について説明する。ここでは、活性度として尤度を計算する。ステップ60での計算の手順を図4に示す。

【0051】テストアミノ酸配列におけるシステイン位置数を  $n_p$ 、アライメントされたテストデータ数を  $m$ 、さらに、2つのシステイン位置の  $i$  番目の組み合わせを  $D_i$  とする。ステップ20で仮定した確率的規則の構造から、組み合わせ  $D_i$  が SS結合を構成する確率  $P(S | D_i)$  は、次式で計算すると仮定する。

【0052】

$$P(S | D_i) = \prod_{j=1}^{m-1} P(S | v_{ij})$$

20

【0053】さらに、この  $i$  番目の組み合わせにおいて、 $j$  番目のアミノ酸から  $j+1$  番目のアミノ酸への変化が表1の  $k$  番目のアミノ酸変化であるとするば、

【0054】

$$p(S | v_{ij}) = \hat{p}(k)$$

【0055】とする。

30 【0056】

$$\hat{p}(k)$$

【0057】は、既にステップ20で学習されている。

【0058】例えば、次に示すように、図4に示すようなシステイン位置1, 2のテストデータが得られたとする。

CCNCNCNNN

CNCCNCNN

ただし、ここで、Nは、システインではないアミノ酸を表す。この位置1, 2の組み合わせからは、8個のアミノ酸種類の変化が左から右に順番に得られる。すなわち、CC→CN、CN→CC、CC→NC、NC→C、C→CC、CC→NN、NN→NC、NC→NN、NN→NNであり、表1の番号において、2、5、3、9、4、15、12、16に対応する。すると、この組み合わせの尤度は、次のように計算される。

【0059】

$$\hat{p}(2) \times \hat{p}(5) \times \hat{p}(3) \times \hat{p}(9) \times \hat{p}(4) \times \hat{p}(15) \times \hat{p}(12) \times \hat{p}(16)$$

50 【0060】ステップ70では、すべてのシステインの

組み合わせに対して、ステップ60で計算した尤度から、SS結合を構成する組み合わせの候補を選出する。ここでは、テストアミノ酸配列に含まれるSS結合の数が $N_{SS}$ とした場合に、例えば、ステップ60で求めた尤度の大きい順に $N_{SS}$ 個の組み合わせを選ぶ。ただし、テストアミノ酸配列すべてのシステインがSS結合を構成する場合（すなわち、 $n_p = 2 * N_{SS}$ ）は、尤度の大きい順に $N_{SS} - 1$ 個の組み合わせを選ぶ。

【0061】ステップ80では、ステップ70で選出されたSS結合を構成する組み合わせの候補に矛盾がないかどうかを検査する。もし、候補の中に矛盾がなければ、選出された組み合わせ候補を予測値としてステップ100へ進む。ただし、テストアミノ酸配列すべてのシステインがSS結合を構成する場合（すなわち、 $n_p = 2 * N_{SS}$ ）に選出された候補に矛盾がなければ、それら候補の組み合わせと残りの一つの組み合わせを予測値としてステップ100へ進む。もし、選出された候補に矛盾がある場合は、ステップ90へと進む。

【0062】ステップ90では、 $N_{SS}$ 個の組み合わせ全体の尤度が最も大きくなるように、組み合わせ最適化を行なう。ここでは、例えば、対称相互結合型ニューラルネットに対して、ボルツマン・アニーリングを用いることにより、最適な組み合わせを求める方法について述べる。

【0063】まず、各組み合わせにネットワークのニューロンユニットを割り当て、ユニット $i$ の出力値を $x_i$ （ただし、 $0 \leq x_i \leq 1$ ）とする。

【0064】ここで、ネットワークのエネルギーを次式で与える。

【0065】

$$E = \sum_i \sum_{j \in I(i)} w_{ij} x_i x_j - \sum_i \theta_i x_i$$

$$w_{ij} = \begin{cases} K & \text{I番目の組み合わせとJ番目の組み合わせが矛盾する時} \\ 0 & \text{I番目の組み合わせとJ番目の組み合わせが矛盾しない時} \end{cases}$$

【0066】ただし、

【0067】

$$K = \max_i \left| \theta_i \left( i = 1, \dots, \frac{n_p(n_p-1)}{2} \right) \right|$$

$\theta_i$ :  $i$ 番目の組み合わせの尤度に対するしきい値

【0068】とする。

【0069】次に、上記エネルギー $E$ が最小になる $x_i$ の組を求める方法について述べる。ここでは、 $x_i$ は0または1とする。まず、 $x_i$ に適当に0または1を割り当てる。次に、以下の手順を適当な停止条件が満たされるまで反復する。

【0070】 $n_p (n_p - 1) / 2$ 個のユニットの中で、1個（例えば、 $i$ 番目）のユニットを選び、 $\Delta E = (2x_i - 1) (-\sum w_{ij} x_j + \theta_i)$ 、 $P = [1 + \tanh(-\Delta E / T(t))] / 2$ を計算する。得られた $P$ を用い、 $P : 1 - P$ の割合でランダムに1:0を発生し、1ならば、 $x_i = 1 - x_i$ とする。このときの $x_i$ により計算されるネットワークのエネルギー値 $E$ が、これまでに求められた $E$ の中で最小であれば、 $x_i$ を出力値として保存しておく。

10 【0071】停止条件としては、一定回数の繰り返し後終了、 $E$ が一定値を下回ると終了、等が考えられる。また、 $T(t)$ としては、

【0072】

$$\frac{\text{Const}}{\ln(t+2)} (\text{Const:定数})$$

【0073】等が考えられる。停止後、出力値が1となった組み合わせの中で、 $\theta_i$ の大きい順に、 $N_{SS}$ 個の組み合わせを選び、これを予測値としてステップ100へ進む。

20 【0074】以上の組み合わせ最適化は、ボルツマン・アニーリングに限らず、他の組み合わせ最適化手法も使用可能である。

【0075】ステップ100では、ステップ80またはステップ90で得られたSS結合を構成するシステイン位置の組み合わせの予測値を、本予測手法の予測結果として出力する。

【0076】以上の実施例は、SS結合に限らず、遠距離相互作用からなるすべての化学結合に対して適用可能である。

30 【0077】

【発明の効果】以上説明したように、本発明によれば、遠距離相互作用からなる化学結合、特に、SS結合に関し、システインを含む立体構造未知のテストアミノ酸配列が与えられたときに、SS結合を構成するようなシステインの組み合わせを高い精度で予測することが可能である。例えば、アミノ酸配列上の相同性が25%以下であるトリプシンとエラスターゼをデータとして使用し、エラスターゼのSS結合の組み合わせを学習し、トリプシンのすべてのSS結合の組み合わせを正しく予測することができる。

【図面の簡単な説明】

【図1】本発明の一実施例を示すフローチャートである。

【図2】図1中の学習データ生成方法およびテストデータ生成方法を示す模式図である。

【図3】図1中の予測ステップの手順を示すフローチャートである。

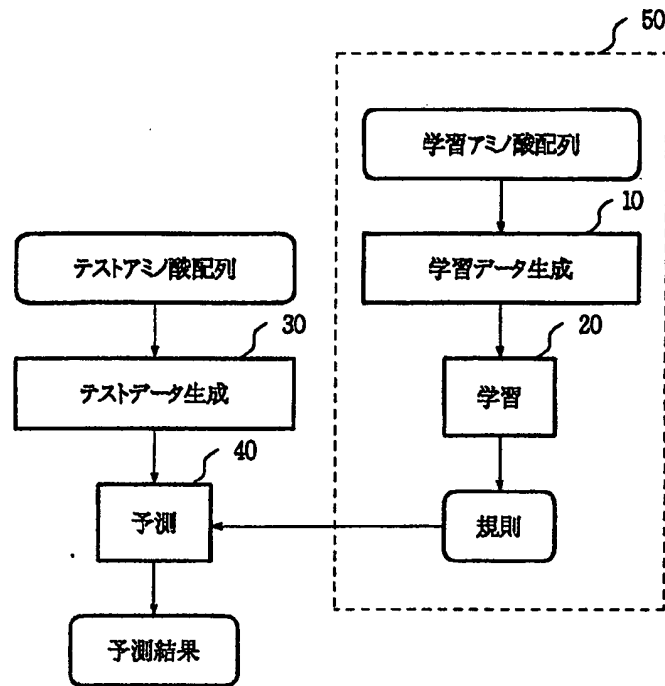
40 【図4】図1の実施例における予測方法を示す模式図である。

## 【符号の説明】

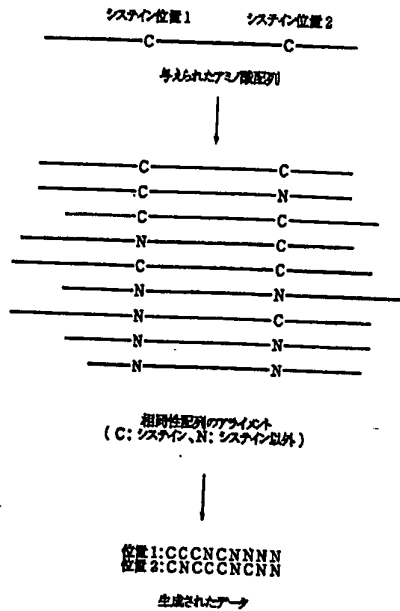
10 学習データ生成  
20 学習  
30 テストデータ生成  
40 予測  
50 学習

60 活性度計算  
70 組み合わせ候補選出  
80 組み合わせ内矛盾検査  
90 組み合わせ最適化  
100 予測結果出力

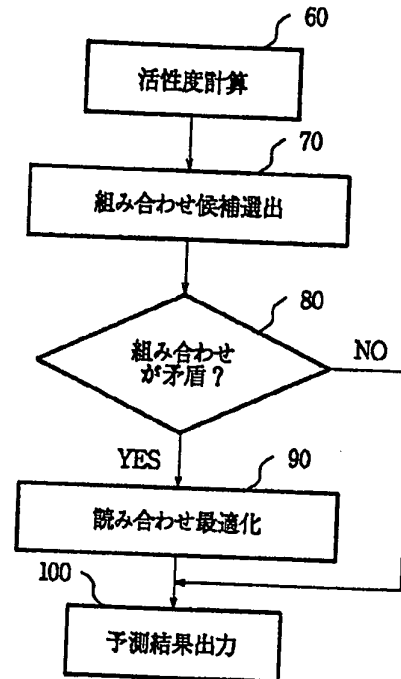
【図1】



【図2】



【図3】



【図4】

